

NASSP Bulletin

<http://bul.sagepub.com>

Uses and Misuses of Standardized Tests

W. James Popham

NASSP Bulletin 2001; 85; 24

DOI: 10.1177/019263650108562204

The online version of this article can be found at:

<http://bul.sagepub.com/cgi/content/abstract/85/622/24>

Published by:

 SAGE Publications

<http://www.sagepublications.com>

On behalf of:

[National Association of Secondary School Principals](#)

Additional services and information for *NASSP Bulletin* can be found at:

Email Alerts: <http://bul.sagepub.com/cgi/alerts>

Subscriptions: <http://bul.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Uses and Misuses of Standardized Tests

W. James Popham

Examines five tests by three publishers currently used in high schools today, and discusses four appropriate and inappropriate uses of these tests. Asserts that assessment literacy on behalf of educators is essential in order to avoid the misuse of standardized tests.

The Law of the Hammer reflects a well-known truth that if you give a hammer to a child, the child will soon identify an enormous number of things in need of hammering. American educators most likely believe that there is an analogous Law of the Standardized Achievement Test because such tests are currently being used for an almost unlimited number of purposes. If an important educational decision needs to be made, and test data are considered relevant, then it is more than likely that someone will try to rely on the results of standardized achievement tests. But tests, as is true with hammers, can sometimes hit the wrong targets.

Focusing on Five Tests

In the following analysis, I want to consider the appropriate as well as the inappropriate uses of standardized achievement tests. I refer specifically to those educational tests that are designed to assess students' skills and knowledge in particular subject fields and are to be administered and interpreted in a standard, predetermined manner. I realize that there are nationally standardized achievement tests as well as state-specific standardized achievement tests being used by educators these days. Those state-level tests, how-

Note. Adapted from a presentation given at the annual meeting of the American Educational Research Association, New Orleans, La., April 24–28, 2000.

W. James Popham is professor emeritus of the Graduate School of Education and Information Studies at the University of California, Los Angeles. Over the years he has been an occasional contributor to the *Bulletin*, beginning with an article in the December 1961 issue and most recently with an article in the December 2000 issue. He lives on the island of Kauai in Hawaii. Correspondence concerning this article may be sent to wpopham@ucla.edu.

ever, are not only linked to a particular state's curriculum, but they can vary substantially in the way they are constructed.

Accordingly, I am going to focus only on the five nationally standardized tests now widely used in the nation's schools, namely, the *California Achievement Tests* and *Comprehensive Tests of Basic Skills* (published by CTB/McGraw-Hill), the *Iowa Tests of Basic Skills* (Riverside), and the *Metropolitan Achievement Tests* and *Stanford Achievement Tests* (Harcourt Educational Measurement).

Only three publishers create and distribute the five nationally standardized achievement tests. Because I do not want to be accused of sniping at any one of these three test-publishing firms, I will simply refer to them as the Big Three when I cite something that has been written by one of the three about the proper or improper use of their own tests. I will be dealing only with *achievement* tests, not *aptitude* tests. Aptitude tests are intended to predict a test taker's likely success in a future situation, typically a future academic setting such as when the SAT or ACT are used to predict how successfully high school students will perform after they get to college. The proper and improper use of aptitude tests is an important, but a different, topic.

Appropriate Uses of Standardized Achievement Tests

There are uses of standardized achievement tests that I regard as acceptable. Taken together, these four uses justify the existence of standardized achievement tests. Hence, I am not opposed to such tests. I think they should exist and even flourish, but I do not think they should be misused. There are four uses of standardized achievement tests that are altogether appropriate.

Informing Parents About Their Children's Relative Achievements

One important use of nationally standardized achievement tests is to give parents an idea about how their children rate, in various subjects, against the performance of a national comparison group. The norm-referenced results yielded by such tests can be quite useful to Marty's mom and dad who discover that Marty scores at the 89th percentile in reading but only at the 34th percentile in mathematics.

From their earliest beginnings, standardized achievement tests have been conceptualized and constructed so that they provide accurate and fine-grained discriminations among test takers. The discriminating efficiency of such tests provides parents with a meaningful fix on how their child compares with a national norm group in responding to a test's items in, for example, language arts or social studies. Such comparative interpretations "enable home and school to work together in the student's best interests."¹

¹ Big Three Member, *Educators' Guide*, current edition.

Informing Teachers About Their Students' Relative Achievements

In much the same way that a parent can gain insights about a child's relative standing in different subject areas, so too can teachers profit from identifying their own students' comparative strengths and weaknesses. A fifth-grade teacher who discovers that Lee scores relatively high in science but relatively low in math can begin to think how to add to Lee's math insights, thereby subtracting from Lee's math weaknesses.

I believe that this identification and use of students' relative strengths and weaknesses ought to be limited either to the subject areas themselves (such as science or reading) or to two or three fairly large chunks of content in a particular subject area. I do not think more fine-grained breakdowns—for example, subskill breakdowns—should be used, for reasons that I will spell out later.

Selecting Students for Special Programs

Because the chief mission of standardized achievement tests is to provide norm-referenced comparisons among students, such tests are ideally suited for informing decisions in fixed-quota settings in which there are more applicants than openings for a special program. The program might be an enrichment activity for gifted children or a remedial activity for low-performing children. Because standardized achievement tests can accurately identify who performs best or worst in responding to a set of items, these tests are quite appropriate whenever students must be chosen from an excessively large applicant pool.

Allocating Supplemental Resources

A fourth appropriate role for standardized achievement test arises when state or district educational policymakers are trying to decide how to distribute discretionary resources such as special funds for additional staff development support or, perhaps, supplemental monies to finance after-school tutorial sessions for low-performing students.

It seems to me that, other things being equal, it makes more sense to assign such supplemental resources to the schools (or, in the case of a state, to the districts) where students' performances on standardized achievement tests indicate that serious instructional action must be taken. I am not suggesting that the schools whose students perform poorly on standardized achievement tests are instructionally ineffective. On the contrary, the students in those schools may be receiving first-rate instruction but come to school from homes that fail to provide the academic support so often linked to students' high test scores. Whatever the cause (and it surely might be poor teaching), the allocation of supplementary resources to low-performing

schools seems more sensible than spreading finite supplemental resources so thinly that neither high-performing nor low-performing schools benefit.

If used in one or more of these four ways, standardized achievement tests may be educationally useful assessment tools. I am sure that there are other suitable uses for standardized achievement tests, but the four I have identified here are, in my view, the most important uses to which such tests should be put.

Inappropriate Uses of Standardized Achievement Tests

I have also identified four inappropriate uses of standardized achievement tests to match the four appropriate uses just treated. Because I think the misuses of standardized tests are particularly serious, I have strong opinions and concerns about these misuses.

Evaluating Schools

Because so many people, including many educators, believe that schools should be judged chiefly on the basis of students' scores on standardized achievement tests, I will address this misuse first. I certainly concur that a school staff's effectiveness should be primarily evaluated on the basis of students' measured achievements. But those achievements should not be measured by standardized achievement tests. Yet the publishers of standardized achievement tests, although dispensing politically correct rhetoric about the need for other evidence of school quality, continue to tout their tests as meaningful contributors to the evaluation of educational effectiveness:

[A] standardized achievement test can provide valuable information about the progress of individuals and groups and the *effectiveness of educational programs* [*italics added*].²

Educators need to remember what the primary purpose of standardized achievement tests really is, namely, to detect sufficient differences among test takers so that sensitive norm-referenced comparisons can be made. To pull off that mission, the developers of such tests sometimes include items apt to be answered correctly by students who either (a) come from advantaged socioeconomic backgrounds or (b) were fortunate enough to inherit above-average academic aptitudes such as verbal or quantitative capacities. Test items based chiefly on these two factors almost always produce the desired spread of student scores because both socioeconomic status and inherited academic aptitudes reflect what children bring to school, not what they learn there.

² Big Three Member, *Teacher's Guide*, current edition.

In addition, the publishers of standardized achievement tests must, for marketing purposes, base their tests on the content that they believe best represents a common-denominator “national curriculum.” Thus, these test publishers must create a one-size-fits-all test covering the content most apt to be emphasized throughout the nation. There is, however, a high likelihood that the curricular preferences of educators in a given locale will not be well aligned with the content of a standardized achievement test.

A school’s educators ought to collect test-based evidence that students are making meaningful achievement gains, but other assessments should also be used to judge success. Standardized achievement tests should not be used to evaluate schools, and every educator needs to understand why.³

Evaluating Teachers

If standardized achievement tests ought not to be used to evaluate a group of teachers in a school, it is surely unsound to use such tests to appraise an individual teacher. Anyone familiar with schools realizes that the caliber of a given teacher’s students can vary dramatically from year to year. To evaluate teachers based on year-to-year changes in their students’ performances on a standardized achievement test is downright laughable, given the potential for shifting ability levels of a teacher’s students. This year’s group of exceptionally advanced students may be replaced next year by students with more learning problems or inadequate preparation. Yet some naïve educational policymakers are advocating a teacher evaluation model that simply subtracts last year’s students’ average test scores from this year’s students’ average test scores. Simple it is; smart it is not.

Fortunately, the publishers of standardized achievement tests recognize this misuse of their product, and some publishers even go on record to discourage it. After pointing out a series of difficulties with judging teacher effectiveness on the basis of standardized achievement tests, one publisher notes, “It is strongly recommended that student test results not be used as a criterion for evaluating teacher performance.”⁴

Promoting or Grading Students

A third inappropriate use of standardized achievement tests arises when teachers use results of such tests to grade students or when test scores are used to promote or retain students. Because of the probable mismatch

³ As might be guessed, I have railed more than a few times about this particular misuse of standardized achievement tests. Educators will find a more extensive treatment of the issue in Popham, W. J. 2000. *Modern educational measurement: Practical guidelines for educational leaders*. 3rd ed. Boston: Allyn & Bacon. Parents will find the misuse of tests treated in Popham, W. J. 2000. *Testing! Testing! What every parent should know about schools tests*. Boston: Allyn & Bacon.

⁴ *Big Three Member, Planning Guide*, current edition.

between a standardized achievement test's "national" content and a particular teacher's instructionally emphasized content, use of scores from such tests for grading purposes is unwise. Given the timing of a standardized achievement test's administration and the delay in the return of students' scores, use of these tests for grading purposes is fairly rare. But whenever it does occur, it is wrong.

Similarly, test publishers point out that standardized achievement tests sample student knowledge and skills at any grade level and "are not intended to be 'end of course' tests."⁵ As one test publisher points out, "If a retention decision is to be made, classroom assessment data gathered by the teacher over a period of months is likely to form a highly relevant and accurate basis for making such a decision."⁶

Making Classroom Instructional Decisions

A fourth and final misuse of standardized achievement tests occurs when teachers try to use the tests or students' test scores to make day-to-day classroom decisions. One publisher argues that a test's content should not be used "to decide which instructional objectives should be taught at a given grade level" because the test's questions constitute only a small sample of the potential questions that might be asked.⁷ Yet a page before this in its printed materials, the same publisher extols the instructional dividends obtainable from one of the publisher's standardized achievement tests.

The following are some of the specific purposes the publisher contends that the tests were designed to serve:

1. To help teachers determine the extent to which individual students in their classes have the knowledge and skills needed to deal successfully with the academic aspects of an instructional program the teacher has planned;
2. To estimate the general developmental level of students so that teachers may adapt materials and instructional procedures to meet individual needs;
3. To identify each student's areas of greatest and least development to use in planning individual instructional goals and approaches;
4. To establish a baseline of achievement information so that the monitoring of year-to-year developmental changes may begin.⁸

⁵ Big Three Member, *Planning Guide*, current edition.

⁶ Big Three Member, *Educators' Guide*, current edition.

⁷ Big Three Member, *Educators' Guide*, current edition.

⁸ Big Three Member, *Educators' Guide*, current edition.

However, the chief obstacle to using standardized achievement tests for purposes of teachers' instructional decision making is insurmountable. Put simply, these tests do not provide teachers with sufficiently clear descriptions of the knowledge and skills represented by the test's items. If the descriptions of the test-represented knowledge and skills are not sufficiently clear for teachers to focus their instruction in an attempt to promote students' mastery of such knowledge and skills, then the teacher's instruction is almost certain to be off target.

Following is a description of one reading objective, along with its corresponding subskills, found in a widely used standardized achievement test. This description is all that a teacher receives in the way of descriptive information regarding the objective:

Evaluate and Extend Meaning

Subskills: generalize; fact/opinion; author—purpose/point of view/tone/bias; predict/hypothesize; extend/apply meaning; critical assessment.⁹

Although those subskills give teachers a somewhat better idea of what is being tested than the unadorned objective does, there is still considerable slack in the description of the test-represented content for this objective. Suppose a teacher decides to focus in on one of the subskills, "generalize." What does this subskill really mean? How can teachers plan lessons if they do not understand what they are supposed to be teaching? Is it any wonder that some teachers, dismayed by such assessment obfuscation, teach to a test's actual items?

Those who claim that classroom teachers can get an accurate idea of their students' progress in objective mastery or, worse, in subskills mastery, have not counted the items per objective or items per subskill in most standardized achievement tests. One publisher calculates subskill scores using as few as three items per subskill. One can only hope that all three items are exceptionally strong and indicative. After all, they are being asked to carry out a serious psychometric mission.

Despite reams of promotional rhetoric from publishers about the instructional dividends derivative from their standardized achievement tests, the use of such tests to make classroom instructional decisions is inappropriate. Are standardized achievement tests more helpful to a teacher's instructional decision making than no tests at all? Yes, but not tremendously more helpful.

Only one of the four misuses of standardized achievement tests—evaluating schools—is so widespread that I think emergency action is required.

⁹ Big Three Member, *Planning Guide*, current edition.

Making classroom instructional decisions on the basis of standardized achievement test scores is also a serious misuse of tests; however, because most teachers cannot figure out how to derive meaningful instructional insights from standardized achievement tests or from students' scores on those tests, this is a fairly feckless misuse. Hence, the evaluation of schools by educators, the lay public, and politicians on the basis of students' standardized achievement test scores needs priority attention.

Medicine for Measurement Misuse

Advances in the pharmaceutical field mean that it is now possible to get one or more medications for a great many ailments that, in earlier years, were essentially untreatable. But what sort of medicine should educators look for if they hope to counter the effects of the inappropriate use of standardized achievement tests and to increase the incidence of the appropriate use of these assessment devices? My answer is quite simple: Every relevant constituency's assessment literacy must be bolstered. I refer not only to educators—heaven knows they need it—but also to educational policymakers, media representatives, and citizens.

Today's educational tests are far too important to be misused by those who are supposed to use them properly. Today's educational tests are far too important to allow such misuses to continue. Although a healthy dose of assessment literacy, all by itself, will not work educational miracles, it is a sensible place to begin treatment. 🦋